Fundamental Frequency and Model Order Estimation Using Spatial Filtering

Karimian-Azari, Sam; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)

DOI (link to publication from Publisher):
10.1109/ICASSP.2014.6854748

Publication date:
2014

Document Version
Early version, also known as pre-print

FUNDAMENTAL FREQUENCY AND MODEL ORDER ESTIMATION USING SPATIAL FILTERING

Sam Karimian-Azari, Jesper Rindom Jensen and Mads Græsbøll Christensen

Audio Analysis Lab, AD:MT, Aalborg University, email: {ska, jrj, mgc}@create.aau.dk

ABSTRACT

In signal processing applications of harmonic-structured signals, estimates of the fundamental frequency and number of harmonics are often necessary. In real scenarios, a desired signal is contaminated by different levels of noise and interferers, which complicate the estimation of the signal parameters. In this paper, we present an estimation procedure for harmonic-structured signals in situations with strong interference using spatial filtering, or beamforming. We jointly estimate the fundamental frequency and the constrained model order through the output of the beamformers. Besides that, we extend this procedure to account for inharmonicity using unconstrained model order estimation. The simulations show that beamforming improves the performance of the joint estimates of fundamental frequency and the number of harmonics in low signal to interference (SIR) levels, and an experiment on a trumpet signal show the applicability on real signals.

***Index Terms*—** Harmonic signal, pitch estimation, model order estimation, microphone arrays, frequency-domain beamforming.

1. INTRODUCTION

In real life, we often have multiple signal sources present at the same time, which has a detrimental impact on the quality and intelligibility of a recorded speech signal. We can improve the quality of a desired signal by choosing an appropriate enhancement method, which can be categorized in three different groups: statistical, filtering, and subspace methods [1]. In the enhancement of harmonic-structured signals as considered here, e.g., voiced speech, both the fundamental frequency and number of harmonics estimates are necessary in filter designs (for example [2–4]). Therefore, we require to estimate these parameters. The estimation of the fundamental frequency, or pitch in audio signal processing, is a challenging problem with applications in enhancement, separation, classification, compression, etc., and different methods have been investigated in the single-channel case [1, 5]. The estimation of number of harmonics is another problem in enhancement of harmonic-structured signals. This integer-valued parameter relating to the number of sinusoidal components must be estimated from the received signals to yield accurate pitch estimates and high-quality enhancement, and some methods have been investigated in the single-channel case [6].

In most of the state-of-the-art methods for fundamental frequency and number of harmonics estimations, the desired signal is assumed to be degraded by additive white Gaussian noise [7–10]. For example, the Markov-like weighted least-squares (WLS) [11] (see also [1,12]) and the maximum a posteriori (MAP) [6,13] methods are fundamental frequency and number of harmonics estimators for only one signal source. In a situation with the presence of interference having the harmonic structure, which is very common, some methods are available to estimate the parameters of multiple signal sources [1]. In these methods, the basic assumption is that the desired signal has higher power than the interferers [2, 14], something that is not always the case. Besides that, multiple harmonic-structured signals with spectral overlap may cause a wrong estimate of the fundamental frequency and the number of harmonics. Furthermore, the inharmonicity problem [15], which is the phenomenon that the frequencies of the harmonics are not exact integers of a fundamental frequency, results in a model mismatch, and leads to biased parameter estimates, e.g., in stiff-stringed instruments.

Exploiting spatial separation is a solution to separate multiple signals using multiple microphones, and beamforming is one such technique to estimate the signal arriving from the desired direction [16] using different source localization methods which have been investigated in [17]. In this paper, we estimate both the fundamental frequency and the number of harmonics, which we call the model order, of a harmonic-structured signal using a beamforming technique to separate the desired signal from high power interferers, which are spatially separated, e.g., by using broadband minimum variance distortionless response (MVDR) [18, 19] beamforming. We can also estimate the model order of the desired signal from the output of the beamformer [20] using the MAP method with the constrained harmonic-model, consisting of a fundamental frequency and its integers. Because of the problem of inharmonicity and harmonic frequencies mismatch, we extend this method for the unconstrained model, consisting of independent sinusoidal components, to estimate both the fundamental frequency and model order. Then the fundamental frequency estimate will be performed using the WLS

This work was funded by the Villum Foundation.

method [11].

The rest of this paper is organized as follows. In Section 2, we introduce the multi-source signal model that the work is based on and apply it in beamforming. In Section 3, we derive the constrained and unconstrained model order and fundamental frequency estimates, and then explore the results of simulations in Section 4. In closing, the work is discussed in Section 5 along with its relation to state-of-the-art.

2. PROBLEM FORMULATION

2.1. Signal model

We consider N independent sources of harmonic acoustic waves, which are placed at different spatial positions, that propagate acoustic waves from their respective direction of arrival (DOA), i.e., θ_n for $n = 1, \dots, N$, relative to a receiver. We assume a microphone array with a set of M omnidirectional microphones receives these acoustic waves besides random noise, i.e., $y_m(t)$ and $v_m(t)$ for $m = 1, \dots, M$. Then, we model the combination of harmonic-structured signal sources, i.e., $x_n(t) = \sum_{l=1}^{L_n} \alpha_{n,l}\, e^{\jmath(l\omega_n t + \varphi_{n,l})}$ that ω_n is the fundamental frequency with L_n number of harmonics with the magnitude of $\alpha_{n,l}$ and phase of $\varphi_{n,l}$,

$$y_m(t) = \sum_{n=1}^{N} \sum_{l=1}^{L_n} \alpha_{n,l}\, e^{\jmath(l\omega_n t + \varphi_{n,l})} e^{-\jmath l \omega_n \Delta\tau_{m,n}} + v_m(t), \quad (1)$$

where $\jmath = \sqrt{-1}$, and $\Delta\tau_{m,n}$ is the time difference of arrival between the mth and the first microphone for the nth source. By expressing the signal model (1) in the frequency-domain vector notation [19], the received broadband signals $\mathbf{Y}(\omega) = [\, Y_1(\omega) \dots Y_M(\omega) \,]^{\mathrm{T}}$ are formulated as functions of the steering vector $\mathbf{d}(\theta_n, \omega)$, signal sources $X_n(\omega)$, and noise $\mathbf{V}(\omega)$, defined similar to $\mathbf{Y}(\omega)$, as

$$\mathbf{Y}(\omega) = \sum_{n=1}^{N} \mathbf{d}(\theta_n, \omega) X_n(\omega) + \mathbf{V}(\omega), \quad (2)$$

where the steering vector is the set of phase shifts between microphones defined at each subband by choosing the first microphone as the reference

$$\mathbf{d}(\theta_n, \omega) = [\, 1 \;\; e^{-\jmath\omega\Delta\tau_{2,n}} \;\; \cdots \;\; e^{-\jmath\omega\Delta\tau_{M,n}} \,]^{\mathrm{T}}. \quad (3)$$

With the aim of the spatial source separation, we can write the spatial correlation matrix, by the assumption of uncorrelated signal sources and noise, as

$$\begin{aligned} \mathbf{R}_{\mathbf{Y}}(\omega) &= \mathrm{E}\{\mathbf{Y}(\omega)\mathbf{Y}^{\mathrm{H}}(\omega)\} \\ &= \sum_{n=1}^{N} \mathbf{d}(\theta_n, \omega) J_{X_n}(\omega) \mathbf{d}^{\mathrm{H}}(\theta_n, \omega) + \mathbf{R}_{\mathbf{V}}(\omega), \end{aligned} \quad (4)$$

where $\mathrm{E}\{\cdot\}$ denotes mathematical expectation, and the superscript $^{\mathrm{H}}$ the transpose-conjugate operator. We define $J_{X_n}(\omega) = \mathrm{E}\{|X_n(\omega)|^2\}$ as the subband power of each signal source, and the noise correlation matrix as $\mathbf{R}_{\mathbf{V}}(\omega) = \mathrm{E}\{\mathbf{V}(\omega)\mathbf{V}^{\mathrm{H}}(\omega)\}$.

2.2. Spatial filtering

All the complex values of the microphone outputs at the subband ω are applied to a complex-valued spatial filter $\mathbf{H}(\theta, \omega)$, or a beamformer as we refer to it, of the length M at each candidate direction θ subject to $\mathbf{H}^{\mathrm{H}}(\theta, \omega)\mathbf{d}(\theta, \omega) = 1$. In general, the output signal will be

$$Z(\theta, \omega) = \mathbf{H}^{\mathrm{H}}(\theta, \omega)\mathbf{Y}(\omega), \quad (5)$$

and the output power of the designed filters is

$$\begin{aligned} J_Z(\theta, \omega) &= \mathrm{E}\{Z(\theta, \omega) Z^{\mathrm{H}}(\theta, \omega)\} \\ &= \mathbf{H}^{\mathrm{H}}(\theta, \omega)\mathbf{R}_{\mathbf{Y}}(\omega)\mathbf{H}(\theta, \omega). \end{aligned} \quad (6)$$

By considering $X_1(\omega)$ as the desired signal, and substituting (4) into (6) at the direction of the desired signal, i.e., θ_1, we acquire the output power of the beamformer as

$$\begin{aligned} J_Z(\theta_1, \omega) = &\; J_{X_1}(\omega) + \mathbf{H}^{\mathrm{H}}(\theta_1, \omega)\mathbf{R}_{\mathbf{V}}(\omega)\mathbf{H}(\theta_1, \omega) \\ &+ \sum_{n=2}^{N} \mathbf{H}^{\mathrm{H}}(\theta_1, \omega)\mathbf{d}(\theta_n, \omega) J_{X_n}(\omega) \mathbf{d}^{\mathrm{H}}(\theta_n, \omega)\mathbf{H}(\theta_1, \omega) \\ = &\; J_{X_1}(\omega) + \Psi(\theta_1, \omega), \end{aligned} \quad (7)$$

where $\Psi(\theta_1, \omega)$ is a residual noise-plus-interference after filtering. The broadband output power of the filter, and the broadband output power of the noise-plus-interference are, respectively,

$$J_Z(\theta) = \frac{1}{2\pi} \int_0^{2\pi} J_Z(\theta, \omega) \mathrm{d}\omega, \quad (8)$$

$$\Psi(\theta_1) = \frac{1}{2\pi} \int_0^{2\pi} \Psi(\theta_1, \omega) \mathrm{d}\omega = J_Z(\theta_1) - J_{X_1}, \quad (9)$$

where J_{X_1} is the broadband power of the desired signal.

The delay-and-sum (DS) beamformer is designed based on the principle that the directivity pattern of the filter is steered to the DOA of interest, i.e., $\mathbf{H}_{\mathrm{DS}}(\theta, \omega) = \mathbf{d}(\theta, \omega)/M$, and the desired signal can be filtered in the composition of different signals (2) depending on the respective DOA. Besides the directivity pattern criteria, the minimum variance distortionless response (MVDR) beamformer is designed to minimize the output power

$$\min_{\mathbf{H}(\theta, \omega)} \mathbf{H}^{\mathrm{H}}(\theta, \omega)\mathbf{R}_{\mathbf{Y}}(\omega)\mathbf{H}(\theta, \omega) \quad (10)$$
$$\text{s.t.} \quad \mathbf{H}^{\mathrm{H}}(\theta, \omega)\mathbf{d}(\theta, \omega) = 1,$$

then the optimal MVDR filter is given by [21]

$$\mathbf{H}_{\mathrm{MVDR}}(\theta, \omega) = \frac{\mathbf{R}_{\mathbf{Y}}^{-1}(\omega)\mathbf{d}(\theta, \omega)}{\mathbf{d}^{\mathrm{H}}(\theta, \omega)\mathbf{R}_{\mathbf{Y}}^{-1}(\omega)\mathbf{d}(\theta, \omega)}. \quad (11)$$

3. PROPOSED METHOD

The signal source X_n with an integer number of harmonics, i.e., L_n, can be modeled in two ways: the constrained (C) harmonic-model that is the integration of integer frequency coefficients relating to the fundamental frequency ω_n, i.e.,

$$\mathbb{X}_n^{\mathrm{C}}(\omega_n) = [\, X_n(\omega_n)\; X_n(2\omega_n) \ldots X_n(L_n^{\mathrm{C}}\omega_n)\,]^{\mathrm{T}}, \quad (12)$$

and the unconstrained (UC) model that is the integer number of independent periodic signals, i.e.,

$$\mathbb{X}_n^{\mathrm{UC}}(\mathbf{\Omega}_n) = [\, X_n(\omega_{n,1})\; X_n(\omega_{n,2}) \ldots X_n(\omega_{n,L_n^{\mathrm{UC}}})\,]^{\mathrm{T}}, \quad (13)$$

where $\mathbf{\Omega}_n = [\,\omega_{n,1}\;\omega_{n,2}\ldots\omega_{n,L_n^{\mathrm{UC}}}\,]^{\mathrm{T}}$ is a set of unconstrained frequencies. By the assumption of two models, the power of the desired signal can be estimated as

$$J_{X_1}^{\mathrm{C}}(\omega_1) = 2\,\|\,\mathbb{X}_1^{\mathrm{C}}(\omega_1)\,\|_2^2, \quad (14)$$

$$J_{X_1}^{\mathrm{UC}}(\mathbf{\Omega}_1) = 2\,\|\,\mathbb{X}_1^{\mathrm{UC}}(\mathbf{\Omega}_1)\,\|_2^2. \quad (15)$$

We can estimate the model order of a harmonic signal from the output power of a beamformer at the desired direction by minimizing the broadband noise power [20]. For both the constrained and unconstrained models in (12) and (13), we write the broadband output power of the noise-plus-interference from (9) like

$$\Psi^{\mathrm{C}}(\theta_1) = J_Z(\theta_1) - J_{X_1}^{\mathrm{C}}(\omega_1), \quad (16)$$

$$\Psi^{\mathrm{UC}}(\theta_1) = J_Z(\theta_1) - J_{X_1}^{\mathrm{UC}}(\mathbf{\Omega}_1). \quad (17)$$

With the assumption of white Gaussian noise and using N_f frequency samples, we can jointly estimate the fundamental frequency and the number of constrained harmonics using the MAP method in the model order estimation [1, 6] as

$$(\hat{L}_1^{\mathrm{C}}, \hat{\omega}_1^{\mathrm{C}}) \approx \arg\min_{L_1^{\mathrm{C}},\omega_1} \Big\{ N_f \ln[\Psi^{\mathrm{C}}(\theta_1)] + \frac{3}{2}\ln N_f + L_1^{\mathrm{C}} \ln N_f \Big\}, \quad (18)$$

which consists of the log-likelihood function of the noise-plus-interference and the penalty part. The penalty part is estimated through the normalization of the Fisher information matrix for a candidate fundamental frequency and L_1 related amplitudes and phases [14]. We can extend this method for estimating the number of independent sinusoids and the related amplitudes and phases, i.e.,

$$(\hat{L}_1^{\mathrm{UC}}, \hat{\mathbf{\Omega}}_1) \approx \arg\min_{L_1^{\mathrm{UC}},\mathbf{\Omega}_1} \Big\{ N_f \ln[\Psi^{\mathrm{UC}}(\theta_1)] + \frac{5}{2} L_1^{\mathrm{UC}} \ln N_f \Big\}. \quad (19)$$

To estimate the fundamental frequency that has the best match to the frequency estimates obtained using the unconstrained model, i.e., $\hat{\mathbf{\Omega}}_1$, we apply the WLS method [11]:

$$\hat{\omega}_1^{\mathrm{UC}} = \frac{\sum_{l=1}^{\hat{L}^{\mathrm{UC}}} l\, |X_1(\omega_{1,l})|^2\, \omega_{1,l}}{\sum_{l=1}^{\hat{L}^{\mathrm{UC}}} l^2\, |X_1(\omega_{1,l})|^2}. \quad (20)$$

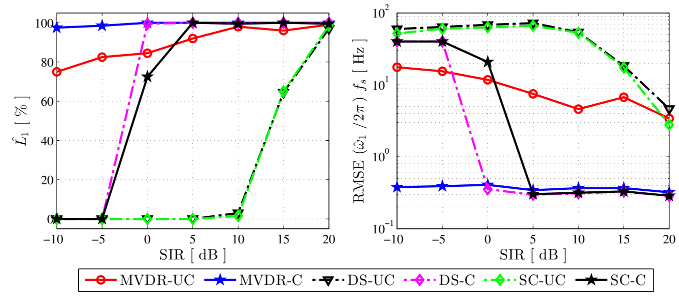

Fig. 1. Performance of the model order and the fundamental frequency estimators for different SIRs [dB], with SNR = 20 dB, and $\Delta\omega_n/2\pi$ = 0.00025.

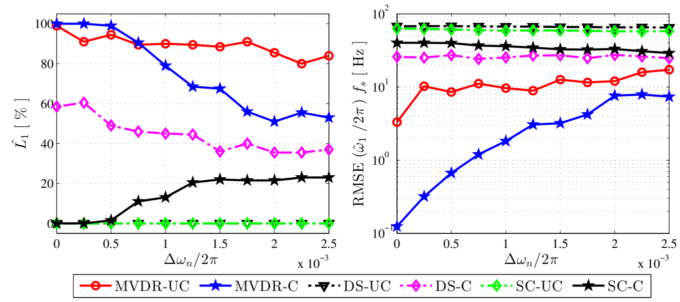

Fig. 2. Performance of the model order and the fundamental frequency estimators for different $\Delta\omega_n/2\pi$, with SNR = 20 dB, and SIR = -1.5 dB.

4. SIMULATION RESULTS

In the following, we evaluate the proposed method and compare the results with single-channel (SC) results in different experiments using synthetic data, and also in a simulation with a real trumpet sound. Then, we measure the root mean square errors (RMSEs) of the fundamental frequency and percentage of correctly model order estimates from 200 Monte-Carlo simulations. In all simulations, we place two synthetic signals at $\theta_1 = 60^\circ$ and $\theta_2 = 40^\circ$, where $\omega_1/2\pi = 0.0225$, $L_1 = 5$ with unit amplitudes, and $\omega_2/2\pi = 0.0275$, $L_2 = 7$, with equal amplitudes depending on signal to interference ratio (SIR) levels, and the sampling frequency is $f_{\mathrm{s}} = 8.0$ kHz. These harmonic-structured signals are simulated like $\mathbf{\Omega}_1 = [\,(\omega_1 + \Delta\omega_{1,1})\;(2\omega_1 + \Delta\omega_{1,2}) \ldots (L_1\omega_n + \Delta\omega_{1,L_1})\,]^{\mathrm{T}}$, where the $\Delta\omega_{1,l}$ is a normal distribution of the frequencies with a variance of zero for simulating the constrained harmonic-model, and a non-zero variance for the unconstrained model with perturbed harmonics.

We model a uniform linear array (ULA) consisting of $M = 10$ omnidirectional microphones, for which the distance between two successive sensors is $\delta = 0.04$ m (smaller than half of the minimum wavelength $\delta \le \lambda_{\min}/2$), and add independent white Gaussian noise to each microphone depending on signal to noise ratio (SNR) levels. The time differences of arrival is $\Delta\tau_{m,n} = (m-1)\delta\sin(\theta_n)/\mathrm{c}$, where the wave propagation speed is assumed to be $c = 343.2$ m/s.

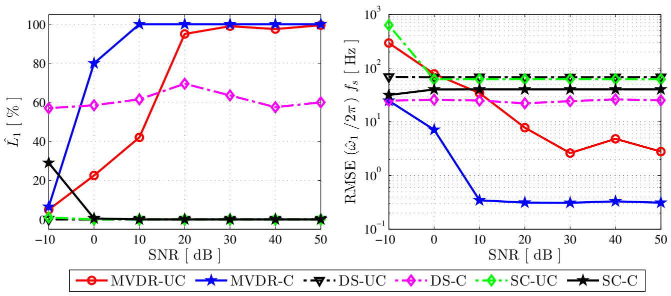

Fig. 3. Performance of the model order and the fundamental frequency estimators for different SNRs [dB], with $\Delta\omega_n/2\pi$ = 0.00025, and SIR = -1.5 dB.

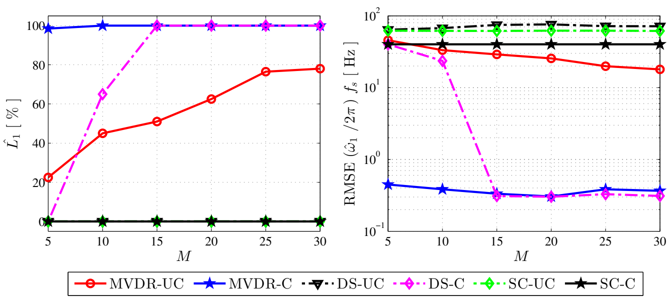

Fig. 4. Performance of the model order and the fundamental frequency estimators using different number of microphones, with $\Delta\omega_n/2\pi$ = 0.00025, SNR = 10 dB, and SIR = -1.5 dB.

The mathematical expectation is estimated by time averaging of B temporal frames [20, 22]. In the MVDR beamforming design (11), the full rank correlation matrix can be guaranteed by choosing $B \geq M$, so that, we choose $B = 30$ in all simulations.

First, the spectral amplitudes of each subband are estimated using a 512 point discrete Fourier transform (DFT). Then, for spectral estimation with large frequency grids, the 65536 point DFT is taken from the zero-padded inverse-DFT of the output signal from the beamformers, and the broadband output power in (18) and (19) are normalized like in [20]. Figure 1 shows that the fundamental frequency and the model order estimation methods are performed in low SIRs using beamforming, and the MVDR beamformer performs better than the DS beamformer. Figure 2 indicates that the unconstrained model order estimate is more accurate in comparison with the constrained harmonic-model in high ranges of perturbed harmonics, $\Delta\omega_n/2\pi \geq 0.001$. The MVDR beamformer outperforms the DS beamformer in low SNRs and number of microphones in figures 3 and 4, respectively. We also conduct an experiment on a trumpet signal with vibrato, as the desired signal, which is corrupted by a synthetic signal similar to in the previous simulations and white Gaussian noise, i.e., SIR = -1.5 dB and SNR = 10 dB. Figure 5 indicates that the unconstrained model order has better estimates than the other model, and the fundamental frequency estimates via the constrained model has better results, compared with the clean signal estimates using the constrained harmonic-model.

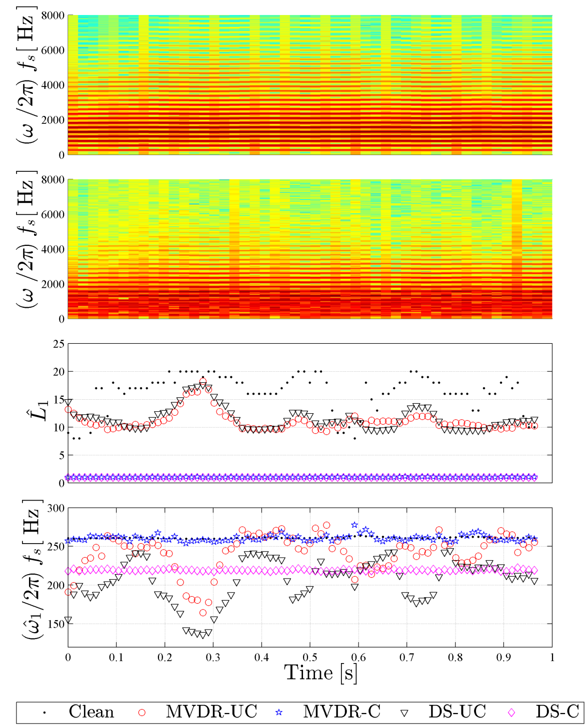

Fig. 5. According to the order of plots from top to down: the spectrogram of a clean trumpet signal $|X_1(\omega)|$, the distorted signal $|Y_1(\omega)|$, the estimates of number of harmonics and the fundamental frequency.

5. DISCUSSION AND CONCLUSION

In this paper, we improve the fundamental frequency and model order estimates of harmonic-structured signals in situations with low SIR. In the multi-channel parameter estimation methods in [10] and [8], it has been considered that a desired signal is contaminated only by Gaussian noise, although in situations with spatially separated interference signals, which are likely in real scenarios, the joint fundamental frequency and constrained model order estimates [14] can be facilitated using spatial filters [20]. Simulations show beamforming will yield better results than the corresponding single-channel estimates, and the optimal MVDR beamformer outperforms the DS, as an example, for closely spaced signal sources. Moreover, through the MAP model order estimation with a uniform probability distribution of random candidates, a general unconstrained model is approached instead of a particular model in [15]. To approach high-resolution of spectral estimates with a minimum variance capability, the DFT method, which we used in our experiments, can be replaced by different methods [5], e.g., unconstrained model extension of the methods in [14] and [23], note that also in the two-dimensional MVDR filter design [23].

6. REFERENCES

[1] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.

[2] M. G. Christensen and A. Jakobsson, "Optimal filter designs for separating and enhancing periodic signals," *IEEE Trans. Signal Process.*, vol. 58, pp. 5969–5983, Dec. 2010.

[3] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, pp. 1124–1138, Oct. 1986.

[4] W. Jin, X. Liu, M. Scordilis, and L. Han, "Speech enhancement using harmonic emphasis and adaptive comb filtering," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18, pp. 356–368, Feb 2010.

[5] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson Education, Inc., 2005.

[6] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, pp. 36–47, Jul. 2004.

[7] J. Tabrikian, S. Dubnov, and Y. Dickalov, "Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model," *IEEE Trans. Speech Audio Process.*, vol. 12, pp. 76 – 87, Jan. 2004.

[8] Z. Zhou, H. So, and M. Christensen, "Parametric modeling for damped sinusoids from multiple channels," *IEEE Trans. Signal Process.*, vol. 61, pp. 3895–3907, Aug 2013.

[9] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Springer-Verlag, 2009.

[10] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 409–412, Mar. 2012.

[11] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Elsevier Signal Process.*, vol. 80(9), pp. 1937–1944, 2000.

[12] M. G. Christensen, P. Vera-Candeas, S. D. Somasundaram, and A. Jakobsson, "Robust subspace-based fundamental frequency estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 101–104, Mar. 2008.

[13] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, pp. 2726 –2735, Oct. 1998.

[14] M. G. Christensen, J. L. Højvang, A. Jakobsson, and S. H. Jensen, "Joint fundamental frequency and order estimation using optimal filtering," *EURASIP J. on Applied Signal Processing*, vol. 2011, pp. 1–18, Jun. 2011.

[15] T. D. Rossing, F. R. Moore, and P. A. Wheeler, *The Science of Sound*. Addison Wesley, 3 ed., 2002.

[16] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.

[17] M. S. Brandstein and H. Silverman, "A practical methodology for speech source localization with microphone arrays," *Comput. Speech Language*, 1997.

[18] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[19] J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*, vol. 5. Springer, 2012.

[20] S. Karimian-Azari, J. R. Jensen, and M. G. Christensen, "Fast joint DOA and pitch estimation using a broadband MVDR beamformer," in *Proc. European Signal Processing Conf.*, Sept. 2013.

[21] J. Benesty, Y. Huang, and J. Chen, *Microphone Array Signal Processing*, vol. 1. Springer-Verlag, 2008.

[22] M. E. Lockwood and et al., "Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms," *The Journal of the Acoustical Society of America*, vol. 115, pp. 379–391, Jan. 2004.

[23] A. Jakobsson, S. L. Jr. Marple, and P. Stoica, "Computationally efficient two-dimensional Capon spectrum analysis," *IEEE Trans. Signal Process.*, vol. 48, pp. 2651–2661, Sep. 2000.